第四章 自适应滤波

§4.1 自适应滤波器的基本概念

- 应用维纳滤波器的封闭解和数值解，需要已知：输入信号的**相关矩阵** **R** 和输入信号与期望响应之间的**互相关向量** **p**。

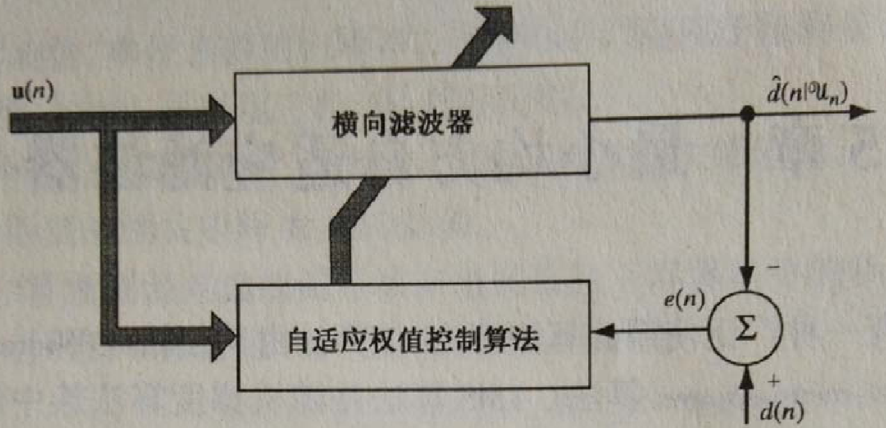
- 实际应用中:

◆(1) 信号的相关矩阵 **R**和互相关向量 **p** 大部分情况下是未知的。

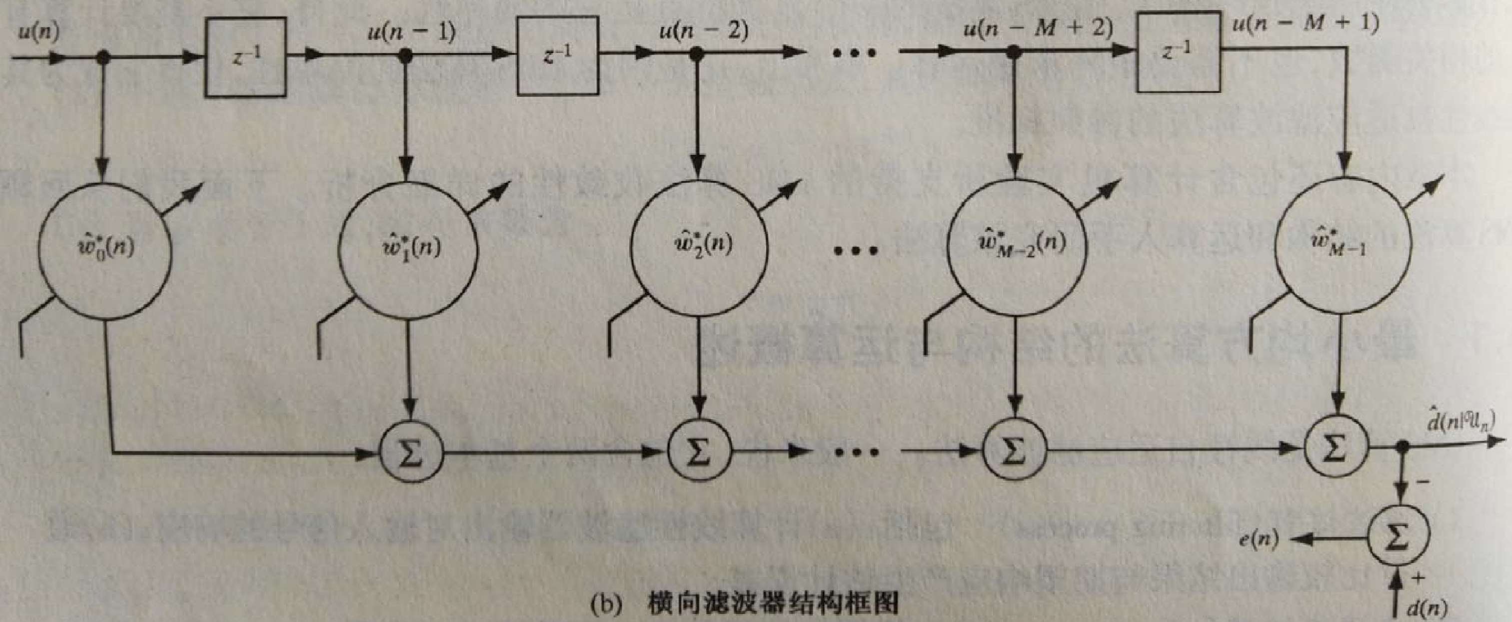
◆(2) 如果信号是时变的，则相关矩阵 **R** 和互相关向量 **p** 是时变的。

- 对于R和 p 是未知，甚至是时变的情形，需要一种能够实时估计或者跟踪信号统计特性的方法，即**自适应滤波**方法。

- 一个自适应滤波器，包含如下三个过程
  - 滤波过程：计算线性滤波器输出对输入信号的响应；
  - 计算估计误差：通过比较输出结果与期望响应产生估计误差。
  - 自适应过程：根据估计误差自动调整滤波器参数。



(a) 自适应横向滤波器框图



(b) 横向滤波器结构框图

$\delta\hat{w}_0(n)$

§4.2 LMS自适应滤波算法

- 前面推导的最速下降算法迭代式为

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu[\mathbf{p}-\mathbf{R}\mathbf{w}(n)]$$

其 中 ， $\mathbf{R}=E[\mathbf{u}(n)\mathbf{u}^H(n)]$, $\mathbf{p}=E[\mathbf{u}(n)d(n)]$

- 为了能够实时处理，可以**采用信号的瞬时值来近似估计**，即

$$\hat{\mathbf{R}}(n)=\mathbf{u}(n)\mathbf{u}^H(n),\ \hat{\mathbf{p}}(n)=\mathbf{u}(n)d(n)$$

- 将上式代入最速下降算法迭代式，可得

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\left[\overbrace{\mathbf{u}(n)d^*(n)}^{\hat{\mathbf{p}}(n)} - \overbrace{\mathbf{u}(n)\mathbf{u}^H(n)}^{\hat{\mathbf{R}}(n)}\mathbf{w}(n)\right]$$

$$= \mathbf{w}(n) + \mu\mathbf{u}(n)\left[d^*(n) - \mathbf{u}^H(n)\mathbf{w}(n)\right]$$

$$= \mathbf{w}(n) + \mu\mathbf{u}(n)\left[d(n) - \overbrace{\mathbf{w}^H(n)\mathbf{u}(n)}^{y(n)}\right]^*$$

$$= \mathbf{w}(n) + \mu\mathbf{u}(n)e^*(n)$$

上式即为**最小均方(LMS)算法**的迭代公式。

最小均方(LMS)算法的迭代公式，包含三个过程：

1） 滤波输出： $y(n) = \mathbf{w}^H(n)\mathbf{u}(n)$

2） 计算估计误差：$e(n)=d(n)-y(n)$

3） 权值向量自适应：$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{u}(n)e^*(n)$

表 5.1　LMS 算法小结

参数：

$M=$ 抽头数(即滤波器长度)

$\mu=$ 步长参数

$$0<\mu<\frac{2}{MS_{\max}}$$

其中 $S_{\max}$是抽头输入 $u(n)$的功率谱密度的最大值，而滤波器长度 M 为中到大

初始化：

如果知道抽头权向量 $\hat{\mathbf{w}}(n)$的先验知识，则用它来选择 $\hat{\mathbf{w}}(0)$的适当值；否则令 $\hat{\mathbf{w}}(0)=\mathbf{0}$

数据：

- 给定的：$\mathbf{u}(n)=n$ 时刻 $M\times 1$ 抽头输入向量

$$=[u(n),u(n-1),\ldots,u(n-M+1)]^T$$

$d(n)=n$ 时刻的期望响应

- 要计算的：$\hat{\mathbf{w}}(n+1)=n+1$ 时刻抽头权向量估计

计算：对 $n=0,1,2\ldots$，计算

$$e(n)=d(n)-\hat{\mathbf{w}}^H(n)\mathbf{u}(n)$$

$$\hat{\mathbf{w}}(n+1)=\hat{\mathbf{w}}(n)+\mu\mathbf{u}(n)e^*(n)$$

§4.3 LMS算法的收敛性

1. LMS算法收敛性证明

- 将LMS算法的迭代公式展开，可得

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n)+\mu[\mathbf{u}(n)e^{*}(n)] \\ &= \mathbf{w}(n)+\mu\{\mathbf{u}(n)[d^{*}(n)-\mathbf{u}^{H}(n)\mathbf{w}(n)]\}\end{aligned}$$

- 上式两边取期望，可得

$$\begin{aligned}E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)]+\mu E[\mathbf{u}(n)d^{*}(n)] \\ -\mu E[\mathbf{u}(n)\mathbf{u}^{H}(n)\mathbf{w}(n)]\end{aligned}$$

- 在步长取值很小的情况下，$\mathbf{w}(n)$ 变化很慢，可以近似为与输入向量 $\mathbf{u}(n)$ 不相关，则近似有

$$E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)] = E[\mathbf{u}(n)\mathbf{u}^H(n)]E[\mathbf{w}(n)]$$

- 于是，迭代式

$$E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)] + \mu E[\mathbf{u}(n)d^*(n)] - \mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]$$

可简化成

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(n)] + \mu\mathbf{p}$$

➢ 设初始权值为 $\mathbf{w}(0)$ ，则有上式递推可得

$$E[\mathbf{w}(1)] = (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(0)] + \mu\mathbf{p}$$

$$\begin{aligned} E[\mathbf{w}(2)] &= (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(1)] + \mu\mathbf{p} \\ &= (\mathbf{I} - \mu\mathbf{R})^2 E[\mathbf{w}(0)] + \mu(\mathbf{I} - \mu\mathbf{R})\mathbf{p} + \mu\mathbf{p} \\ &\vdots \end{aligned}$$

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})^{n+1} E[\mathbf{w}(0)] + \mu\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{R})^i \mathbf{p}$$

- 根据特征值分解，可令 $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$

其中，$\mathbf{Q}$ 为酉矩阵，满足 $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}, \mathbf{Q}^H = \mathbf{Q}^{-1}$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & & \\ & & \vdots & \\ 0 & & & \lambda_M \end{pmatrix}$$

- 则递推式

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})^{n+1} E[\mathbf{w}(0)] + \mu\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{R})^{i}\mathbf{p}$$

可转化为

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{n+1} E[\mathbf{w}(0)]$$

$$+\mu\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{i}\mathbf{p}$$

- 根据酉矩阵的性质 $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$, $\mathbf{Q}^H = \mathbf{Q}^{-1}$, 递推式

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{n+1} E[\mathbf{w}(0)] + \mu\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{i}\,\mathbf{p}$$

等号右边第一项的矩阵可写为

$$\begin{aligned}(\mathbf{I} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{n} &= (\mathbf{Q}\mathbf{Q}^{-1} - \mu\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1})^{n}\\ &= [\mathbf{Q}(\mathbf{I} - \mu\boldsymbol{\Lambda})\mathbf{Q}^{-1}]^{n}\\ &= \mathbf{Q}(\mathbf{I} - \mu\boldsymbol{\Lambda})^{n}\mathbf{Q}^{-1}\end{aligned}$$

- 则递推式可转化为

$$E[\mathbf{w}(n+1)] = \mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})^{n+1}\mathbf{Q}^{-1}E[\mathbf{w}(0)] + \mu\mathbf{Q}\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{\Lambda})^{i}\mathbf{Q}^{-1}\mathbf{p}$$

要使上式收敛，必须满足如下条件：

$$\left|1 - \mu\lambda_{\max}\right| < 1 \quad \Rightarrow \quad 0 < \mu < \frac{2}{\lambda_{\max}}$$

- 条件 $0 < \mu < 2/\lambda_{\max}$ 需要知道相关矩阵的特征值，而实际应用中可能是未知的。
- 由于 $\lambda_{\max} < \sum_{i=0}^{M-1} \lambda_{i+1} = tr[\mathbf{R}] = \sum_{i=0}^{M-1} E[u^2(i)]$，则步长的范围可缩小为

$$0 < \mu < \frac{2}{\sum_{i=1}^{N} E[u_i^2]}$$

其中，$\sum_{i=0}^{M-1} E[u_i^2]$为输入信号的功率，通常可以估计。

◆下面证明，当 $n \to \infty$ 时，$\mathbf{w}(n)$ 收敛于何处。

➢ 当 $n \to \infty$ 时，递推式

$$E[\mathbf{w}(n+1)] = \mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})^{n+1}\mathbf{Q}^{-1}E[\mathbf{w}(0)]$$
$$+\mu\mathbf{Q}\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{\Lambda})^{i}\mathbf{Q}^{-1}\mathbf{p}$$

转化为

$$\lim_{n\to\infty} E[\mathbf{w}(n+1)] = \mu\mathbf{Q}\lim_{n\to\infty}\sum_{i=0}^{n}(\mathbf{I} - \mu\mathbf{\Lambda})^{i}\mathbf{Q}^{-1}\mathbf{p}$$

- 递推式中的矩阵 $\lim_{n\to\infty}\sum_{i=0}^{n}(\mathbf{I}-\mu\mathbf{\Lambda})^i$ 可转化为

$$\lim_{n\to\infty}\sum_{i=0}^{n}(\mathbf{I}-\mu\mathbf{\Lambda})^i=\left[\mathbf{I}-(\mathbf{I}-\mu\mathbf{\Lambda})\right]^{-1}=\frac{1}{\mu}\mathbf{\Lambda}^{-1}$$

将上式代入递推式

$$\lim_{n\to\infty}E[\mathbf{w}(n+1)]=\mu\mathbf{Q}\lim_{n\to\infty}\sum_{i=0}^{n}(\mathbf{I}-\mu\mathbf{\Lambda})^i\mathbf{Q}^{-1}\mathbf{p}$$

可得 $\lim_{n\to\infty}E[\mathbf{w}(n+1)]=\mu\mathbf{Q}\frac{1}{\mu}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1}\mathbf{p}=\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1}\mathbf{p}$

$$=\mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{H}\mathbf{p}=\mathbf{R}^{-1}\mathbf{p}=\mathbf{w}_o$$

2. 步长参数对稳定性和收敛速度的影响

(以实数为例)

- 权值向量为一维的情况

$$\Delta w(n) = w(n) - w_o$$

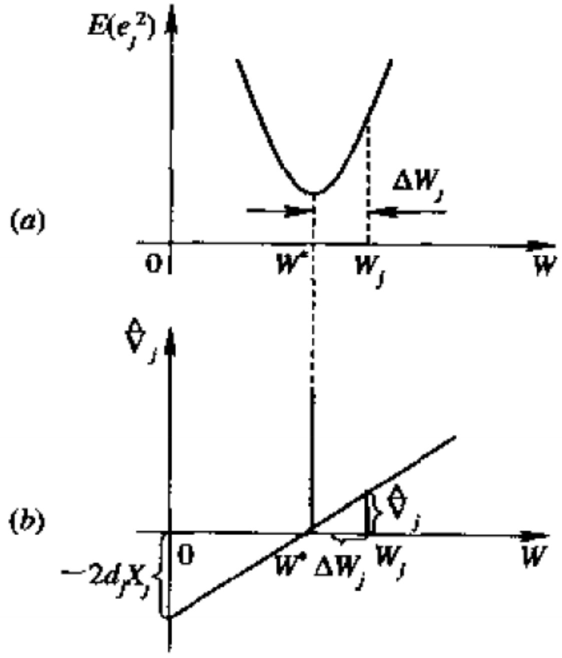


图 4.8　一维情况下的 $E[e_j^2] \sim W$ 及 $\hat{\nabla}_j \sim W$ 关系

➢ 可以证明，$\Delta w(n)$ 越大，则在点 $w(n)$ 处的斜率 $\nabla(n)$ 越大。

➢ 在 n 点处的瞬时估计为

$$\hat{\nabla}(n) = \frac{\partial e^2(n)}{\partial w(n)} = 2e(n)\frac{\partial e(n)}{\partial w(n)}$$

$$= -2d(n)u(n) + 2w(n)u^2(n)$$

➢ 可见，在n点处的瞬时估计

$$\hat{\nabla}(n) = -2d(n)u(n) + 2w(n)u^2(n)$$

与$w(n)$的关系，为一条直线，

其斜率为

$$\frac{\partial \hat{\nabla}(n)}{\partial w(n)} = 2u^2(n)$$

即，$\hat{\nabla}(n)$正比于$w(n)$。

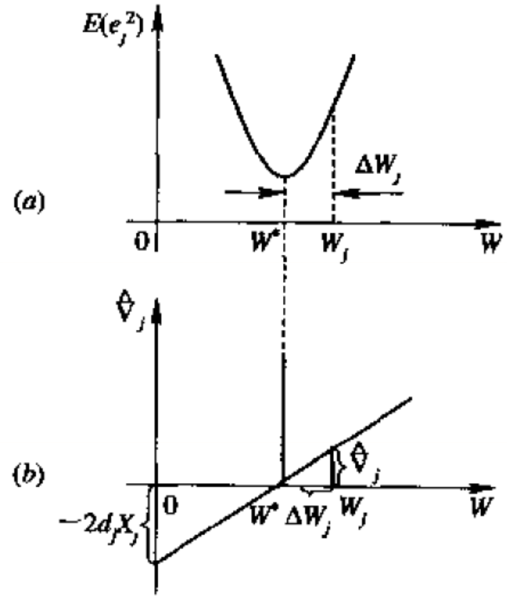


图 4.8 一维情况下的 $E[e_j^2] \sim W$ 及 $\hat{\nabla}_j \sim W$ 关系

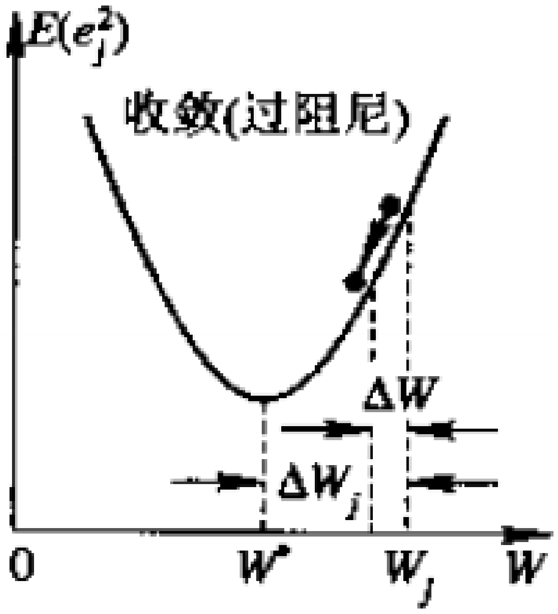
➢ 根据最速下降法公式 $w(n+1) = w(n) - \frac{1}{2}\mu\hat{\nabla}(n)$

每递推一次，$w(n)$ 向 w_o 靠拢的量为

$$\Delta w = w(n) - w(n+1) = \mu u^2(n)\Delta w(n)$$

- 根据 $\Delta w = \mu u^2(n)\Delta w(n)$ ，容易看出

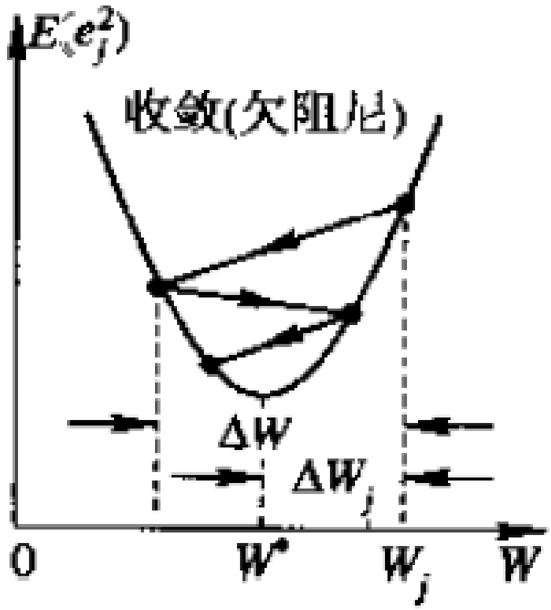
a) 在 $\Delta w < \Delta w(n)$ 的范围内，μ 越大，Δw 越大，收敛越快。此情况称为过阻尼。



(a)

- 根据 $\Delta w = \mu u^2(n)\Delta w(n)$ ，容易看出

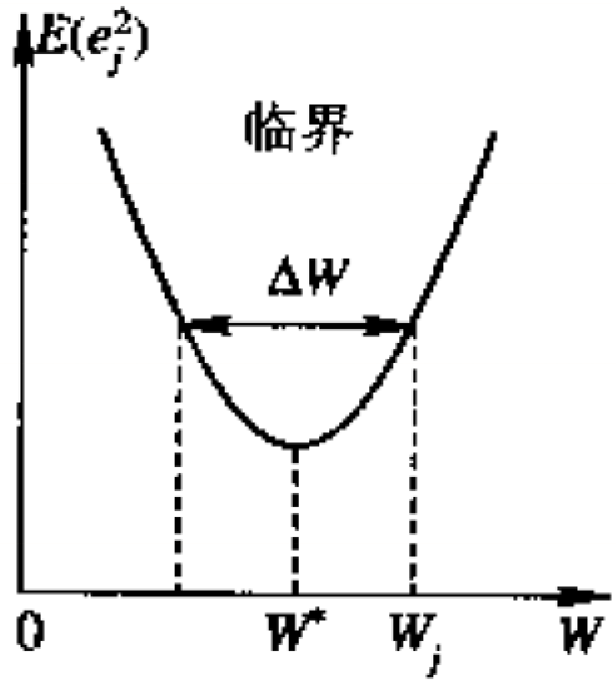
b) μ 加大，但满足 $\Delta w(n) < \Delta w < 2\Delta w(n)$ 的条件，此时将以震荡型轨迹收敛于 w_o 。此情况称为欠阻尼。



(b)

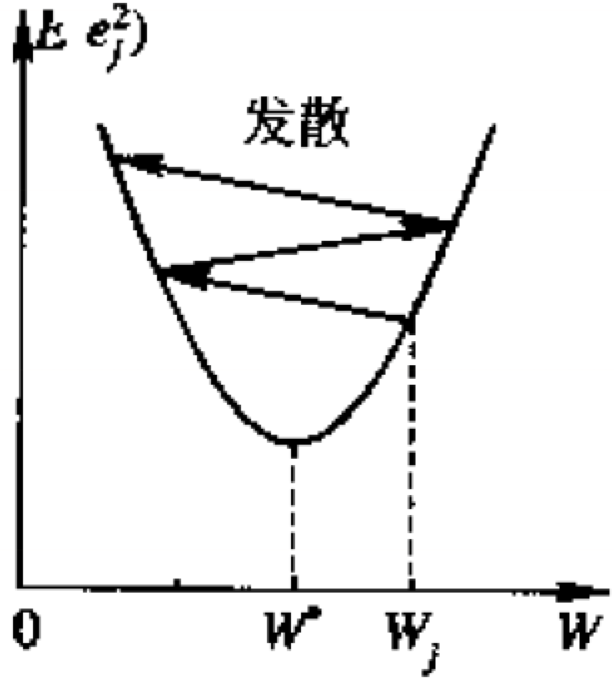
- 根据 $\Delta w = \mu u^2(n)\Delta w(n)$ ，容易看出

c) 当 μ 取值增大，使 $\Delta w = 2\Delta w(n)$ 时，此时的 $w(n)$将不能收敛于 w_o。 此情况称为临界阻尼。



(c)

- 根据 $\Delta w = \mu u^2(n)\Delta w(n)$ ，容易看出
  d) 当 μ 取值过大，使 $\Delta w > 2\Delta w(n)$ 时，此时的 $w(n)$ 将发散，不再收敛于任何值。



(d)

- 通常情况下，为了获得比较小的逼近误差，步长 μ 的值一般取的很小，从而不太会出现(b)(c)(d)的情况。

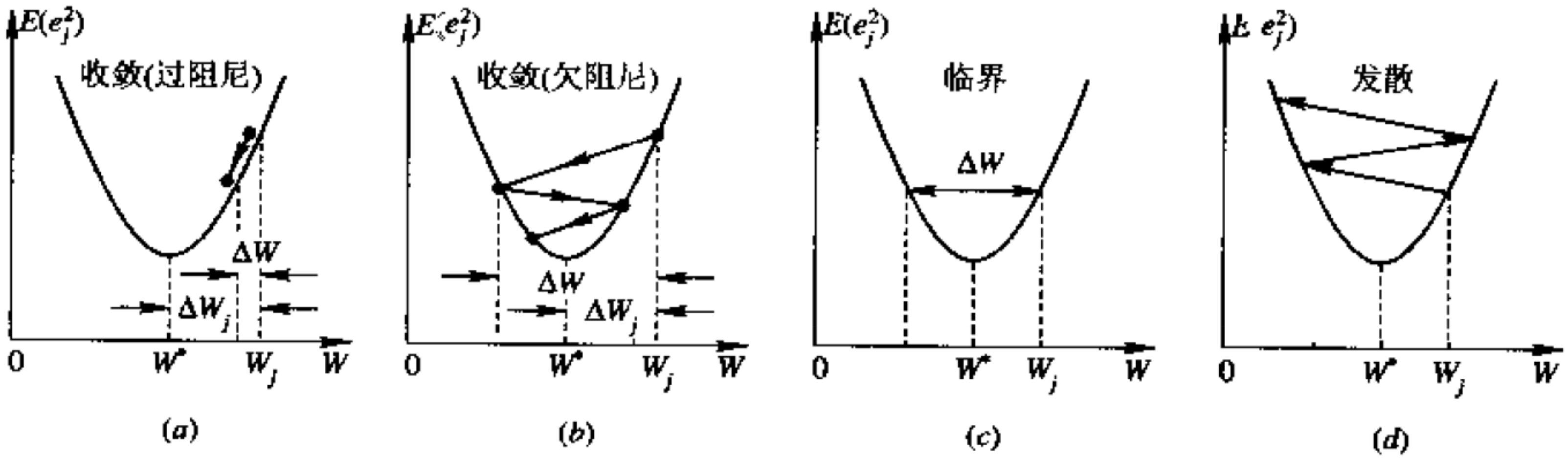
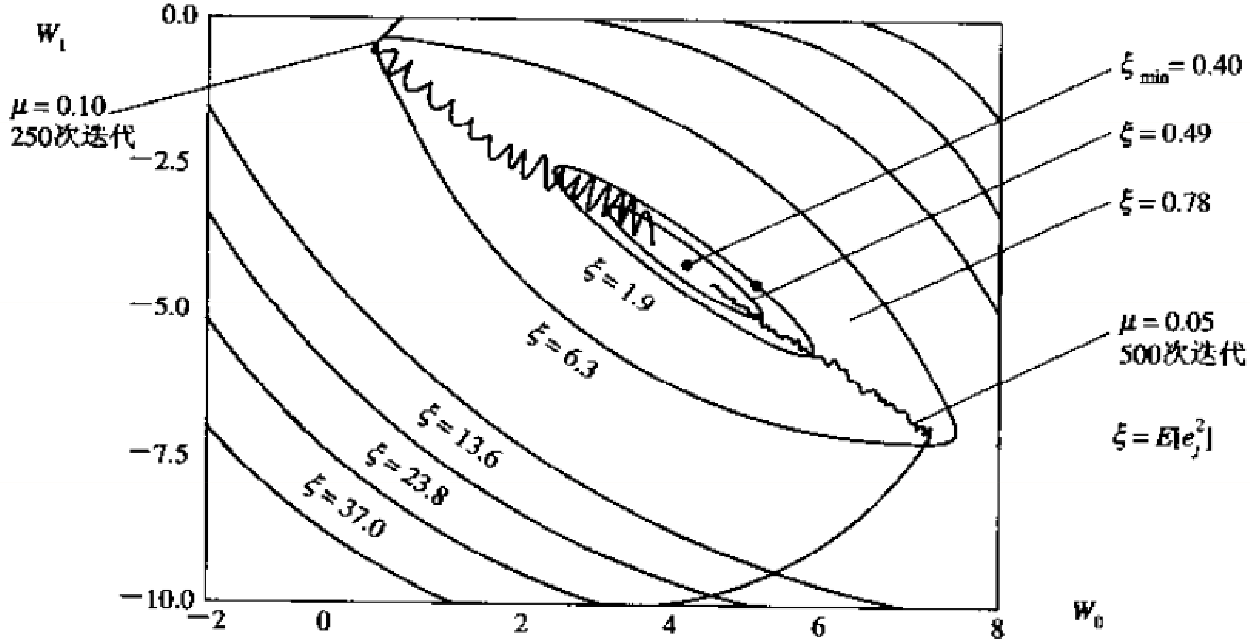


图 4.9 μ的大小对收敛的影响

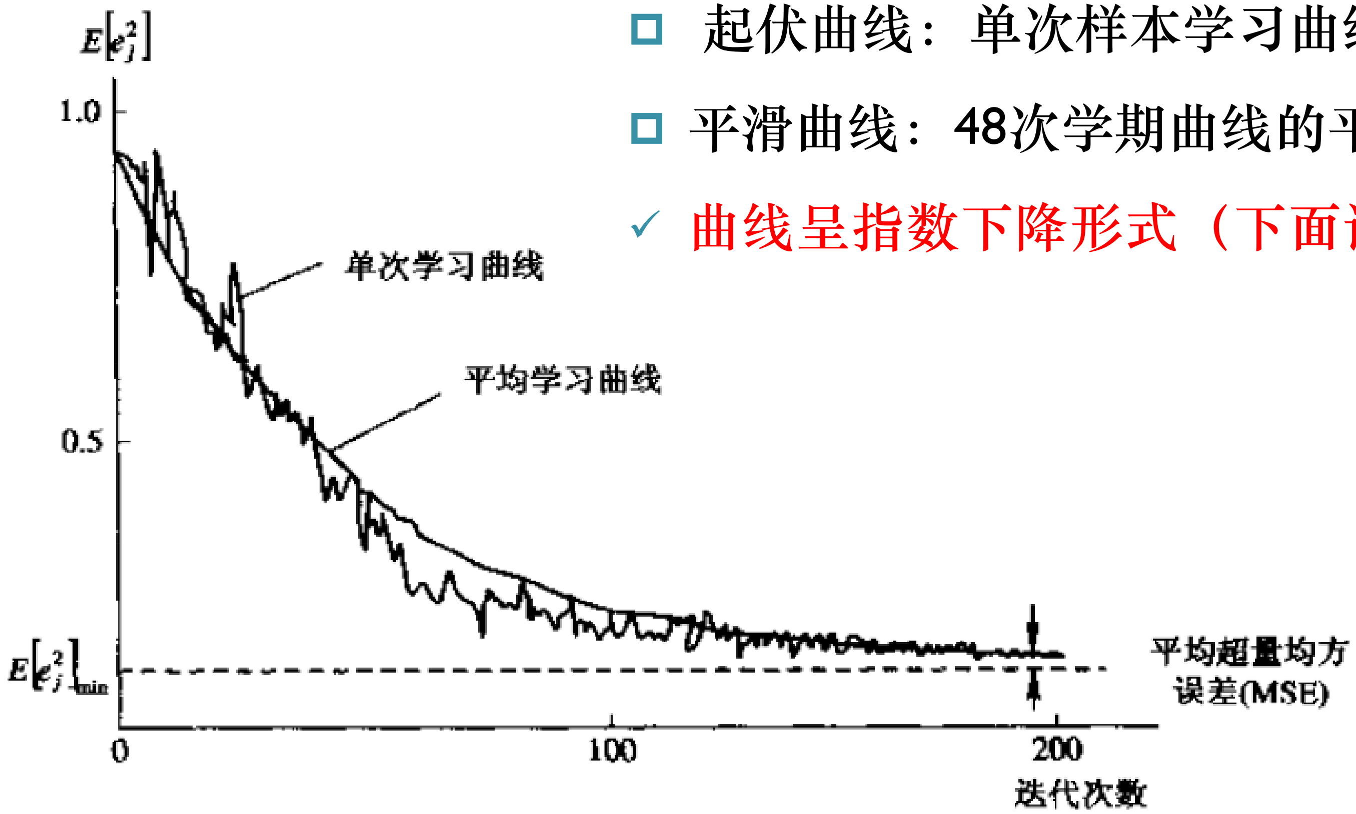
权值向量为二维的情况

- ✓ 大步长($\mu = 0.1$)：收敛速度快，随机抖动性大，收敛后精度低；
- ✓ 小步长($\mu = 0.05$)：较为平稳，精度高，收敛缓慢



3. 自适应算法的学习曲线(以实数为例)

➢ 自适应滤波器的$\mathbf{w}(n)$需要经过一个迭代过程才能达到$\mathbf{w}_o$，即$E[e^2(n)]$趋于$E[e^2(n)]_{\min}$需要一个过程。



- 起伏曲线：单次样本学习曲线
- 平滑曲线：48次学期曲线的平均
- ✓ 曲线呈指数下降形式（下面证明）

- 将维纳解 $\mathbf{w}=\mathbf{w}_o=R^{-1}\mathbf{p}$ 代入均方误差表达式

$$E[e^2(n)]=E[d^2(n)]-2\mathbf{p}^T\mathbf{w}+\mathbf{w}^T\mathbf{R}\mathbf{w}$$

可得

$$\begin{aligned}E[e^2(n)]_{\min}&=E[d^2(n)]-2\mathbf{p}^T\mathbf{w}_o+\mathbf{w}_o^T\mathbf{R}\mathbf{w}_o\\&=E[d^2(n)]-2\mathbf{p}^T\mathbf{w}_o+\mathbf{w}_o^T\mathbf{p}\\&=E[d^2(n)]-\mathbf{w}_o^T\mathbf{p}\end{aligned}$$

则用下式表示上式，可得

$$E[e^2(n)]=E[e^2(n)]_{\min}+\mathbf{w}_o^T\mathbf{p}-2\mathbf{p}^T\mathbf{w}+\mathbf{w}^T\mathbf{R}\mathbf{w}$$

令$\mathbf{v}(n)=\mathbf{w}(n)-\mathbf{w}_o$，进一步化简，可得

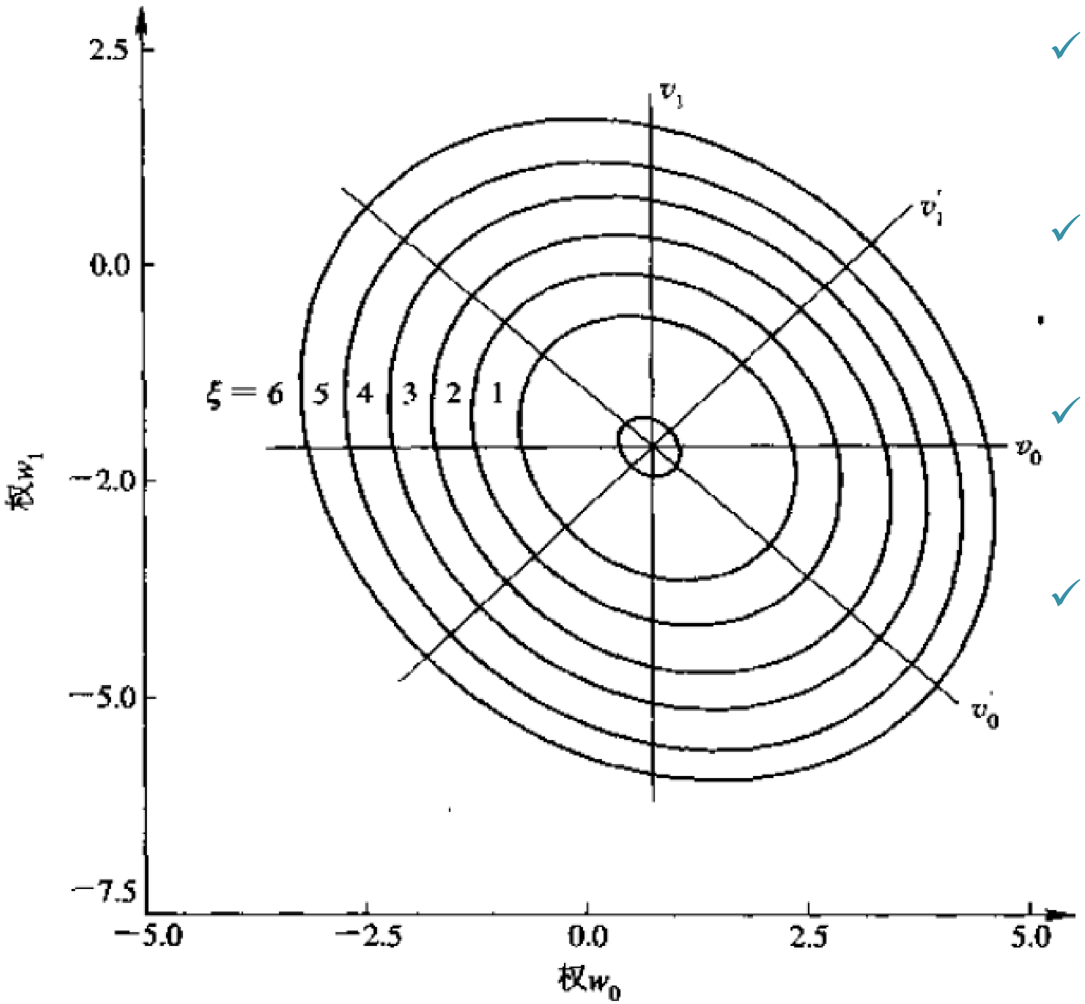
$$
\begin{aligned}
&E[e^2(n)]\\
&=E[e^2(n)]_{\min}+\mathbf{w}_o^T\mathbf{p}-2\mathbf{p}^T\mathbf{w}(n)+\mathbf{w}^T(n)\mathbf{R}\mathbf{w}(n)\\
&=E[e^2(n)]_{\min}+\mathbf{w}_o^T\mathbf{R}\mathbf{w}_o-2\mathbf{w}_o^T\mathbf{R}\mathbf{w}(n)+\mathbf{w}^T(n)\mathbf{R}\mathbf{w}(n)\\
&=E[e^2(n)]_{\min}+\left(\mathbf{w}(n)-\mathbf{w}_o\right)^T\mathbf{R}\left(\mathbf{w}(n)-\mathbf{w}_o\right)\\
&=E[e^2(n)]_{\min}+\mathbf{v}^T(n)\mathbf{R}\mathbf{v}(n)\\
&=E[e^2(n)]_{\min}+\mathbf{v}^T(n)\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{v}(n)
\end{aligned}
$$

根据特征值分解，令 $\mathbf{R}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$，则有

$$
\begin{aligned}
E[e^2(n)] &= E[e^2(n)]_{\min} + \mathbf{v}^T(n)\mathbf{R}\mathbf{v}(n) \\
&= E[e^2(n)]_{\min} + \mathbf{v}^T(n)\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{v}(n) \\
&= E[e^2(n)]_{\min} + \left[\mathbf{Q}^T\mathbf{v}(n)\right]^T \mathbf{\Lambda}\mathbf{Q}^T\mathbf{v}(n) \\
&= E[e^2(n)]_{\min} + \mathbf{v}'^T(n)\mathbf{\Lambda}\mathbf{v}'(n)
\end{aligned}
$$

其中，$\mathbf{v}'(n)=\mathbf{Q}^T\mathbf{v}(n)$。

权值向量为二维情况举例说明几何意义



- 均方误差超抛物面在水平面上的投影；
- 误差椭圆重心不一定在该坐标的原点；
- v坐标系以误差椭圆的重心为原点；
- v′为主轴：

  平移：$\mathbf{v}(n)=\mathbf{w}(n)-\mathbf{w}_o$；

  旋转：$\mathbf{v}'(n)=\mathbf{Q}^T\mathbf{v}(n)$。

求权值误差指数衰减的时间常数

将均方误差 $E[e^2(n)] = E[e^2(n)]_{\min} + \mathbf{v}^T(n)\mathbf{R}\mathbf{v}(n)$

对 $\mathbf{w}(n)$取微分，可得 $\nabla_n = 2\mathbf{R}\mathbf{V}(n)$

根据最速下降法 $\mathbf{w}(n+1) = \mathbf{w}(n) - \dfrac{\mu}{2}\nabla_n$，有

$$\mathbf{w}(n+1) - \mathbf{w}_o = \mathbf{w}(n) - \mathbf{w}_o - \mu\mathbf{R}\mathbf{V}(n)$$

即有 $\mathbf{v}(n+1) = \mathbf{v}(n) - \mu\mathbf{R}\mathbf{v}(n) = \mathbf{Q}[\mathbf{I} - \mu\mathbf{\Lambda}]\mathbf{Q}^T\mathbf{v}(n)$

两边同乘$\mathbf{Q}^T$，上式简化为

$$\mathbf{v}'(n+1) = [\mathbf{I} - \mu\mathbf{\Lambda}]\mathbf{v}'(n)$$

由式 $\mathbf{v}'(n+1)=[\mathbf{I}-\mu\mathbf{\Lambda}]\mathbf{v}'(n)$ 递推，可得

$$\mathbf{v}'(1)=[\mathbf{I}-\mu\mathbf{\Lambda}]^1\mathbf{v}'(0)$$

$$\mathbf{v}'(2)=[\mathbf{I}-\mu\mathbf{\Lambda}]\mathbf{v}'(1)=[\mathbf{I}-\mu\mathbf{\Lambda}]^2\mathbf{v}'(0)$$

$$\vdots$$

$$\mathbf{v}'(n)=[\mathbf{I}-\mu\mathbf{\Lambda}]^n\mathbf{v}'(0)$$

由上式可得如下结论：

- ✓ 每次迭代，等式右边第二项以 $[\mathbf{I}-\mu\mathbf{\Lambda}]$ 速度衰减；
- ✓ 当 $n\to\infty$ 时，$\mathbf{w}(n)=\mathbf{w}_o$；

如果权向量有M个分量，则权值的衰减具有：

- ✓ M个自由度；
- ✓ M个特征值；
- ✓ M个特征振动模式。

第m个振动模式的能量每次衰减为$(1-\mu\lambda_m)$。

定义 $\gamma_m = 1-\mu\lambda_m$，则 γ_m 随迭代次数n作**指数衰减**。

所以 $\gamma_m = e^{-1/\tau_m} = 1-\dfrac{1}{\tau_m}+\dfrac{1}{2!\tau_m^2}+\cdots$

对于比较缓慢而平稳的自适应过程，τ_m一般较大，因此可以略去等式右边的高次项，即

$$\gamma_m = 1-\mu\lambda_m \approx 1-\frac{1}{\tau_m} \quad \Rightarrow \quad \tau_m = \frac{1}{\mu\lambda_m}$$

□均方误差指数衰减的时间常数

因为 $E[e^2(n)] = E[e^2(n)]_{\min} + \mathbf{v}'^T(n)\mathbf{\Lambda}\mathbf{v}'(n)$，则有

$$E[e^2(n)] = E[e^2(n)]_{\min} + \mathbf{v}'^T(0)[\mathbf{I} - \mu\mathbf{\Lambda}]^{2n}\mathbf{v}'(0)$$

由上式可得如下结论：

- ✓ 每次迭代，等式右边第二项以 $[\mathbf{I} - \mu\mathbf{\Lambda}]^2$ 速度衰减；
- ✓ 当 $n \to \infty$ 时，$E[e^2(n)] = E[e^2(n)]_{\min}$；
- ✓ $E[e^2(n)]$ 随 n 增加的衰减比 $\mathbf{w}(n)$ 衰减快一倍。

如果权向量有M个分量，则权值误差均方误差的衰减具有：

- ✓ M个自由度；
- ✓ M个特征值；
- ✓ M个特征振动模式。

定义 $\gamma_m^2=(1-\mu\lambda_m)^2$， 则 $\left(\gamma_m^2\right)^n$ 随迭代次数n作**指数衰减**。

令 $\gamma_m^2=[1-\mu\lambda_m]^2=e^{-1/\tau_{mse}} \Rightarrow \gamma_m=1-\mu\lambda_m=e^{-1/(2\tau_{mse})}$

采用泰勒展开，上式可表示为

$$\gamma_m=I-\mu\lambda_m=e^{-1/(2\tau_{mse})}=1-\frac{1}{2\tau_{mse}}+\frac{1}{2!(2\tau_{mse})^2}+\cdots$$

略去等式右边的高次项，即

$$1-\mu\lambda_m\approx 1-\frac{1}{2\tau_{mse}} \quad \Rightarrow \quad \tau_{mse}=\frac{1}{2\mu\lambda_m}$$

- 随 n 增加，$E[e^2(n)]$ 的衰减比w(n)衰减快一倍。

决定衰减快慢的主要因素

- ✓ 对应于M个权值的自适应滤波器，其权值误差向量的衰减，是由M个自由衰减模式决定；
- ✓ 在最不利的情况下，具有最大衰减常数的分量对衰减起主导作用；
- ✓ 一般情况下的收敛速度由最大时间常数决定，

- 令 **R** 的最小和最大特征值分别为$\lambda_{\min}$ 和 $\lambda_{\max}$ 。
- 如果按照之前的条件$0<\mu<2/\lambda_{\max}$ 取步长的上界$\mu=2/\lambda_{\max}$，则最大的时间常数为

$$\tau_{\max}=\frac{1}{\mu\lambda_{\min}}\geq\frac{1}{2}\frac{\lambda_{\max}}{\lambda_{\min}}$$

- $\lambda_{\max}/\lambda_{\min}$称为输入信号相关矩阵的特征值扩散度。因此，输入信号相关矩阵的特征值扩散度越大，最大的时间常数 $\tau_{\max}$就越大，自适应滤波器收敛就越慢。

- 白噪声对应相关矩阵的特征值扩散度最小，即 $\lambda_{\max}/\lambda_{\min}=1$，因而收敛最快；
- 随着信号的相关性越强，其对应相关矩阵的特征值扩散度越大，因而收敛越慢；

◆ 因此，LMS算法的主要缺点是，当输入信号的相关矩阵特征值较大时，收敛很慢。

✓ **解决思路：将输入信号去相关（白化）**

✓ **主要算法：LMS格型算法、RLS算法、仿射投影算法、子带自适应算法等。**

梯度噪声引起的失调

- 在自适应过程中，若没有环境噪声引起的误差，权向量 $\mathbf{w}(n)$ 最终将收敛于均方性能曲面的最小点 $\mathbf{w}_o$；
- 在此条件下，向量 $\mathbf{v}(n)$的协方差为0，均方误差 $E[e^2(n)]=E[e^2(n)]_{\min}$。

- 实际上，一般只能使 $E[\mathbf{w}(n)] = \mathbf{w}_o$，这是因为LMS算法用瞬时梯度代替梯度时会产生梯度估计误差(梯度噪声)。
- 梯度噪声导致 $\mathbf{w}(n)$ 的值随机起伏。这意味着即使滤波器达到了稳态，均方误差 $E[e^2(n)]$任然大于 $E[e^2(n)]_{\min}$。

- 定义超量均方误差为 $\zeta = E[e^2(n)] - E[e^2(n)]_{\min}$
- 归一化后的超量均方误差，称为失调，即

$$M = \frac{\zeta}{E[e^2(n)]_{\min}}$$

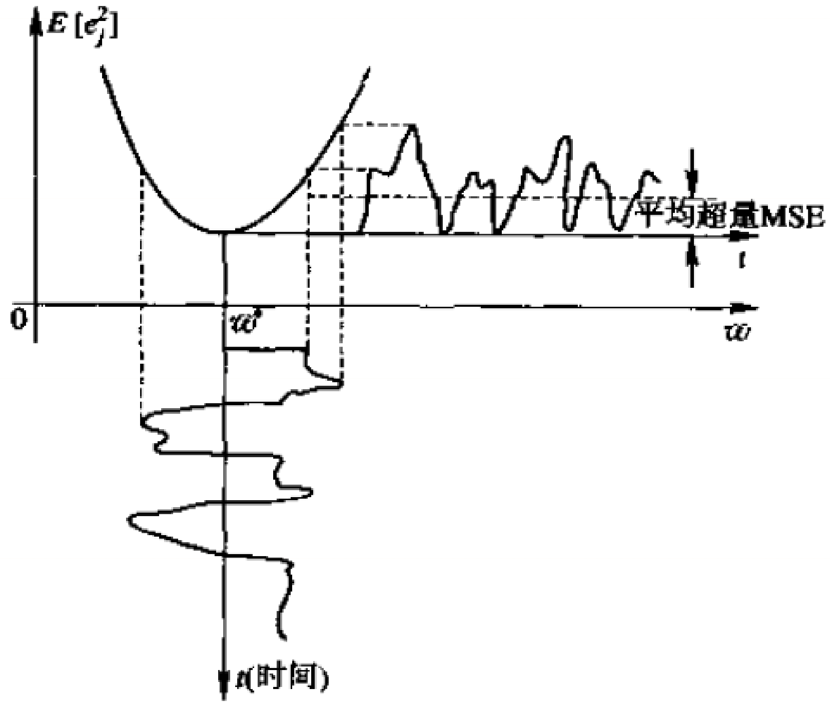


图 4.13 由于梯度噪声而造成的超量均方误差

- 该量可以近似为

$$M = \frac{\mu}{2}\mathrm{tr}[\mathbf{R}]$$

- 失调正比于步长。

§4.4 LMS格型自适应滤波器

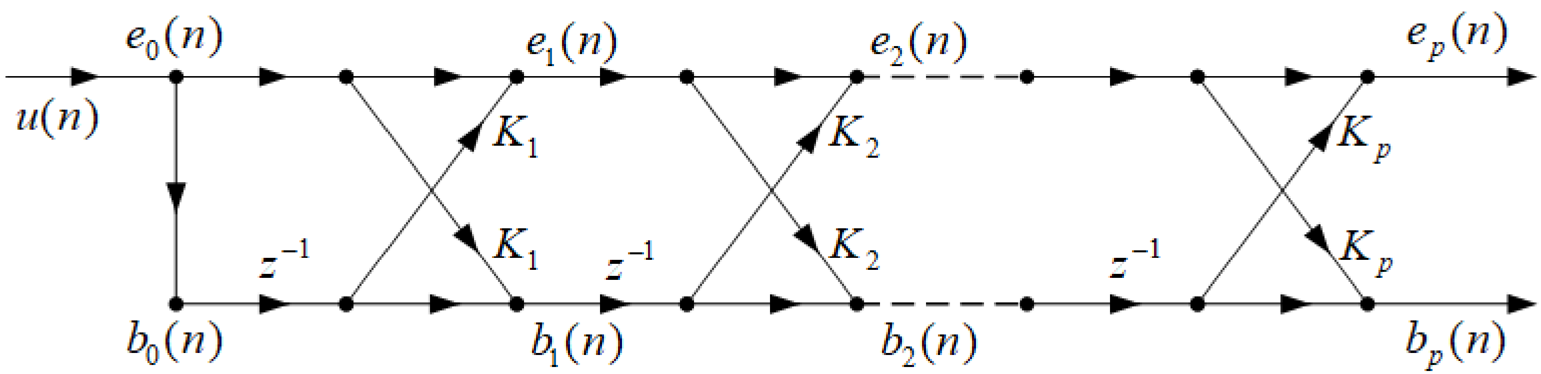
前面介绍的LMS算法，基于横向滤波器。

其优点为：结构简单

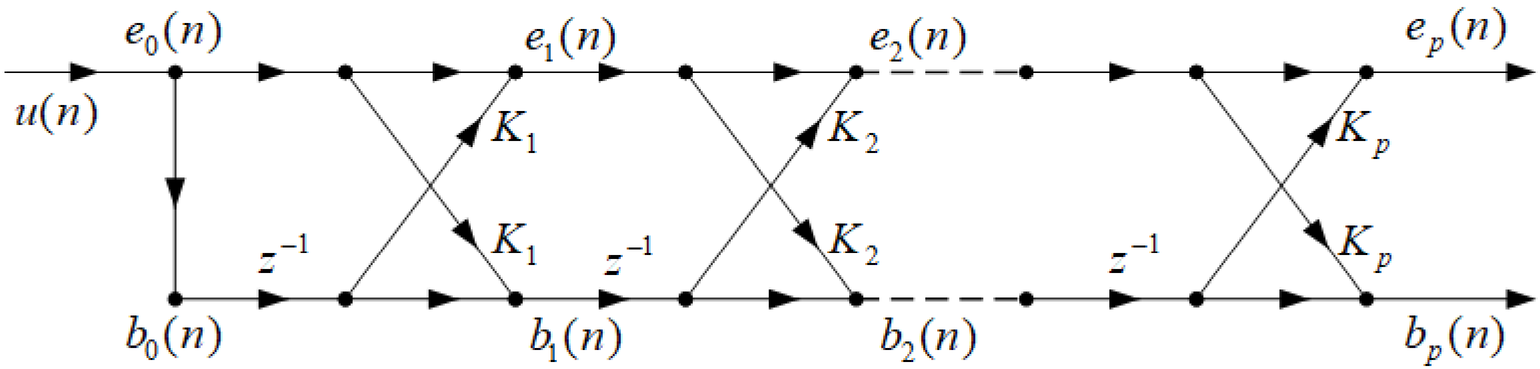
其缺点为：收敛速度慢

克服LMS横向滤波器收敛速度慢的方法之一，

是采用**格型自适应滤波器**。



- P阶格型滤波器，由P个反射系数表示；
- P个反射系数：$K_1, K_2, \cdots, K_P$
- P+1个前项预测误差：$e_0(n), e_1(n), \cdots, e_P(n)$
- p+1个后项预测误差：$b_0(n), b_1(n), \cdots, b_P(n)$



- 由格型滤波器的信号关系,可得:

$$\begin{cases} e_p(n) = e_{p-1}(n) + K_p b_{p-1}(n-1) \\ b_p(n) = b_{p-1}(n-1) + K_p e_{p-1}(n) \end{cases} \quad p \in \{1, 2, \cdots, P\}$$

- 对前项预测误差 $e_p(n)=e_{p-1}(n)+K_p b_{p-1}(n-1)$

取均方值，可得

$$\xi_p \triangleq E[e_p^2(n)] = E\left\{\left[e_{p-1}(n)+K_p b_{p-1}(n-1)\right]^2\right\}$$

$$= E\left[e_{p-1}^2(n)\right] + 2K_p E\left[e_{p-1}(n) b_{p-1}(n-1)\right]$$

$$+ K_p^2 E\left[b_{p-1}^2(n-1)\right]$$

- 如果输入信号是平稳的，那么预测误差也是平稳的，可以证明

$$E[e_p^2(n)] = E[b_p^2(n)] = E[b_p^2(n-1)] = \xi_p$$

- 则前向均方误差可写成

$$\xi_p = \xi_{p-1}(K_p^2 + 1) + 2K_p\Delta_p$$

其中

$$\Delta_p \triangleq E[e_{p-1}(n)b_{p-1}(n-1)]$$

对前向均方误差求导，可得

$$\frac{\partial \xi_p}{\partial K_p} = 2K_p\xi_{p-1} + 2\Delta_p, \qquad p \in \{1, 2, \cdots, P\}$$

当p=1时，$$\frac{\partial \xi_1}{\partial K_1} = 2K_1\xi_0 + 2\Delta_1$$

当p=2时，$$\frac{\partial \xi_2}{\partial K_2} = 2K_2\xi_1 + 2\Delta_2$$

$$\vdots$$

当p=P时，$$\frac{\partial \xi_P}{\partial K_P} = 2K_P\xi_{P-1} + 2\Delta_P$$

其中 $\xi_0 = E[u^2(n)]$, $\Delta_1 = E[u(n)u(n-1)]$

令导数为0，即

$$\frac{\partial \xi_1}{\partial K_1} = 2K_1\xi_0 + 2\Delta_1 = 0$$

$$\frac{\partial \xi_2}{\partial K_2} = 2K_2\xi_1 + 2\Delta_2 = 0$$

$$\vdots$$

$$\frac{\partial \xi_P}{\partial K_P} = 2K_P\xi_{P-1} + 2\Delta_P = 0$$

可得最优的反射系数

$$K_p^o = -\frac{\Delta_p}{\xi_{p-1}} = -\frac{E[e_{p-1}(n)b_{p-1}(n-1)]}{E[e_{p-1}^2(n)]}, \quad p \in \{1, 2, \cdots, P\}$$

递推迭代式

根据最快下降法，每个反射系数的更新公式可写成

$$K_p(n)=K_p(n-1)-\frac{\mu_p}{2}\frac{\partial\xi_p(n-1)}{\partial K_p(n-1)},\quad p\in\{1,2,\cdots,P\}$$

式中，μ_p为第p级的步长(或称为增益)。

容易求得导数

$$\frac{\partial \xi_p(n-1)}{\partial K_p(n-1)} = \frac{\partial e_p^2(n-1)}{\partial K_p(n-1)} = 2e_p(n-1)\frac{\partial e_p(n-1)}{\partial K_p(n-1)}$$

$$= -2e_p(n-1)b_{p-1}(n-2)$$

将上式代入最速下降公式，可得

$$K_p(n) = K_p(n-1) - \mu_p e_p(n-1)b_{p-1}(n-2)$$

步长满足如下条件

$$0 < \mu_p < \frac{1}{\xi_{p-1}}$$

§4.5 RLS自适应滤波算法

- 无论是LMS横向或格型自适应滤波器，在本质上，其收敛速度还是比较慢。
- 为了解决这一问题，可以采用**递推最小二乘**（**RLS**）自适应滤波器。
- **递推最小二乘**（**RLS**）自适应滤波器的理论基础，是著名的最小二乘法。

- 横向自适应滤波器的输出误差为

$$e(n)=d(n)-\mathbf{w}^{T}(n)\mathbf{u}(n)$$

- 建立代价函数

$$\varepsilon(n,\mathbf{w})=\sum_{i=1}^{n}\lambda^{n-i}\left|e(i)\right|^{2}$$

- 求使得上述代价函数最小的权值向量

$$\mathbf{w}(n)=[w_{0}(n),w_{1}(n),\ldots,w_{M-1}(n)]^{T}$$

➢ 为了使$\varepsilon(n,\mathbf{w})=\sum_{i=1}^{n}\lambda^{n-i}\left|e(i)\right|^2$ 最小化，可以令

$$\frac{\partial\varepsilon(n,\mathbf{w})}{\partial\mathbf{w}}=0,$$ 即有

$$\frac{\partial\varepsilon(\mathbf{w}(n))}{\partial\mathbf{w}(n)}=\frac{\partial}{\partial\mathbf{w}(n)}\sum_{i=1}^{n}\lambda^{n-i}[d(i)-\mathbf{w}^T(n)\mathbf{u}(i)]^2$$

$$=-2\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)[d(i)-\mathbf{w}^T(n)\boldsymbol{u}(i)]=0$$

$$=-2\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)[d(i)-\boldsymbol{u}^T(i)\mathbf{w}(n)]=0$$

- 上式可以转化为

$$\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)\mathbf{u}^{T}(i)\mathbf{w}(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)d(i)$$

- 若令

$$\mathbf{R}(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)\mathbf{u}^{T}(i)$$

$$\mathbf{U}(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)\,d(i)$$

- 则有

$$\mathbf{R}(n)\mathbf{w}(n)=\mathbf{U}(n)$$

- 若$\mathbf{R}(n)$为非奇异矩阵，则可求得最优解为

$$\mathbf{w}(n)=\mathbf{R}^{-1}(n)\mathbf{U}(n)$$

- 解 $\mathbf{w}(n)=\mathbf{R}^{-1}(n)\mathbf{U}(n)$ 为滤波器系数的维纳解。下面，来推导其迭代解。
- 利用 $e(n)=d(n)-\mathbf{w}^T(n-1)\mathbf{u}(n)$，可得

$$\mathbf{U}(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)[\mathbf{w}^T(i-1)\mathbf{u}(i)+e(i)]$$

$$=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)[\mathbf{u}^T(i)\mathbf{w}(i-1)+e(i)]$$

$$=\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)\mathbf{u}^T(i)\mathbf{w}(i-1)+\mathbf{u}(i)e(i)]$$

- 将式 $\mathbf{U}(n)=\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)\mathbf{u}^T(i)\mathbf{w}(i-1)+\mathbf{u}(i)e(i)]$

代入 $\mathbf{w}(n)=\mathbf{R}^{-1}(n)\mathbf{U}(n)$ ，可得

$$\mathbf{w}(n)=\mathbf{R}^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)\mathbf{u}^T(i)\mathbf{w}(i-1)]$$

$$+\mathbf{R}^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)e(i)]$$

- 令 $\mathbf{w}_1(n)=\mathbf{R}^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)\mathbf{u}^T(i)\mathbf{w}(i-1)]$

$$\mathbf{w}_2(n)=\mathbf{R}^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)e(i)]$$

前述迭代式可简写为 $\mathbf{w}(n)=\mathbf{w}_1(n)+\mathbf{w}_2(n)$ 。

- 假设自适应滤波器在M个时刻变化不大，即有

$$\mathbf{w}(0)=\mathbf{w}(1)=\cdots=\mathbf{w}(n-1)$$

则可得

$$\begin{aligned}\mathbf{w}_1(n)&=\mathbf{R}^{-1}(n)\left\{\sum_{i=1}^{n}\lambda^{n-i}\left[\mathbf{u}(i)\mathbf{u}^T(i)\right]\right\}\mathbf{w}(n-1)\\&=\mathbf{R}^{-1}(n)\mathbf{R}(n)\mathbf{w}(n-1)\\&=\mathbf{w}(n-1)\end{aligned}$$

- 为了简化 $\mathbf{w}_2(n)=\mathbf{R}^{-1}(n)\sum_{i=1}^{n}\lambda^{n-i}[\mathbf{u}(i)e(i)]$，可以认为遗忘因子 $\lambda=0$ ，这相当于只记住本时刻的结果，而将以前各时刻的结果全部遗忘。

- 采用上述假设之后，可得

$$\mathbf{w}_2(n)=\mathbf{R}^{-1}(n)\mathbf{u}(i)e(i)$$

- 将简化后的 $\mathbf{w}_1(n)$ 和 $\mathbf{w}_2(n)$ 代入原递推式，可得

$$\mathbf{w}(n)=\mathbf{w}(n-1)+\mathbf{R}^{-1}(n)\mathbf{u}(n)e(n)$$

- 递推式 $\mathbf{w}(n)=\mathbf{w}(n-1)+\mathbf{R}^{-1}(n)\mathbf{u}(n)e(n)$ 需要对矩阵 $\mathbf{R}(n)$ 求逆，具有很高的计算量。

- 下面介绍一种递推求解 $\mathbf{R}^{-1}(n)$ 的方法：

根据定义，可将$\mathbf{R}(n)$ 写成

$$\mathbf{R}(n)=\sum_{i=1}^{n}\lambda^{n-i}\mathbf{u}(i)\mathbf{u}^T(i)$$

$$=\left\{\sum_{i=1}^{n-1}\lambda^{n-i}\mathbf{u}(i)\mathbf{u}^T(i)\right\}+\lambda^{n-n}\mathbf{u}(n)\mathbf{u}^T(n)$$

$$=\mathbf{R}(n-1)+\mathbf{u}(n)\mathbf{u}^T(n)$$

- **矩阵求逆引理**

$$(\mathbf{A}+\mathbf{B}\mathbf{C}^T)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{I}-\mathbf{C}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^T\mathbf{A}^{-1}$$

- 利用上述矩阵求逆引理，计算$\mathbf{R}^{-1}(n)$的递推式可变为

$$\mathbf{R}^{-1}(n)=\mathbf{R}^{-1}(n-1)-\frac{\mathbf{R}^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{R}^{-1}(\mathbf{u}-1)}{1+\mathbf{u}^T(n)\mathbf{R}^{-1}(n-1)\boldsymbol{u}(n)}$$

RLS算法的工作步骤

(1) 初始化: $\mathbf{w}(0)=\mathbf{0},\ \mathbf{R}(0)=\mathbf{I}$;

(2) 迭代更新： for n=1到N

$$e(n)=d(n)-\mathbf{w}^T(n)\mathbf{u}(n)$$

$$\mathbf{R}^{-1}(n)=\mathbf{R}^{-1}(n-1)-\frac{\mathbf{R}^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{R}^{-1}(\mathbf{u}-1)}{1+\mathbf{u}^T(n)\mathbf{R}^{-1}(n-1)\boldsymbol{u}(n)}$$

$$\mathbf{w}(n)=\mathbf{w}(n-1)+\mathbf{R}^{-1}(n)\mathbf{u}(n)e(n)$$

➢ RLS算法的改进

移入遗忘因子λ，从而在收敛速度和稳态失调之间进行折中：

$$\mathbf{R}^{-1}(n)=\lambda^{-1}\left[\mathbf{R}^{-1}(n-1)-\frac{\lambda^{-1}\mathbf{R}^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^{T}(n)\mathbf{R}^{-1}(\mathbf{u}-1)}{1+\lambda^{-1}\mathbf{u}^{T}(n)\mathbf{R}^{-1}(n-1)\boldsymbol{u}(n)}\right]$$

移入遗忘因子λ的取值，通常在0.9—1之间。

RLS算法的改进

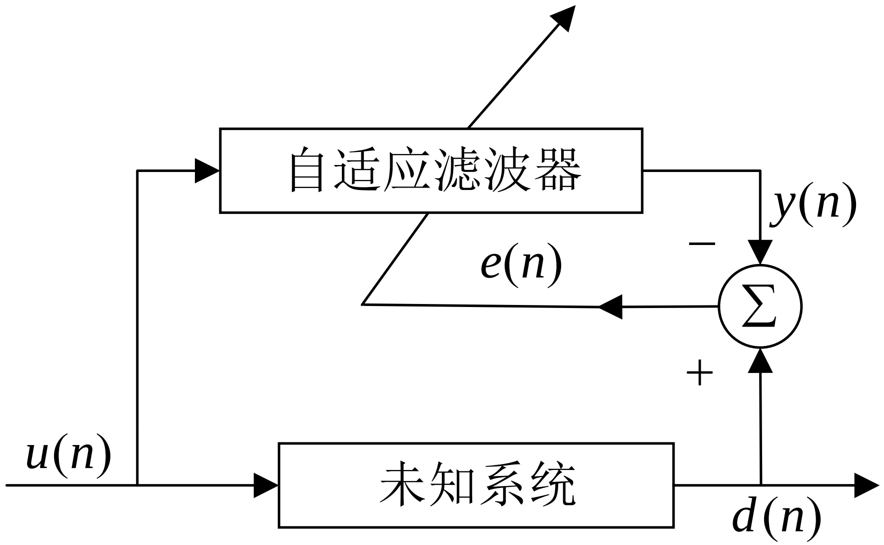
移入遗忘因子λ，从而在收敛速度和稳态失调之间进行折中：

$$\mathbf{R}^{-1}(n)=\lambda^{-1}\left[\mathbf{R}^{-1}(n-1)-\frac{\lambda^{-1}\mathbf{R}^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^{T}(n)\mathbf{R}^{-1}(\mathbf{u}-1)}{1+\lambda^{-1}\mathbf{u}^{T}(n)\mathbf{R}^{-1}(n-1)\boldsymbol{u}(n)}\right]$$

移入遗忘因子λ的取值，通常在0.9—1之间。

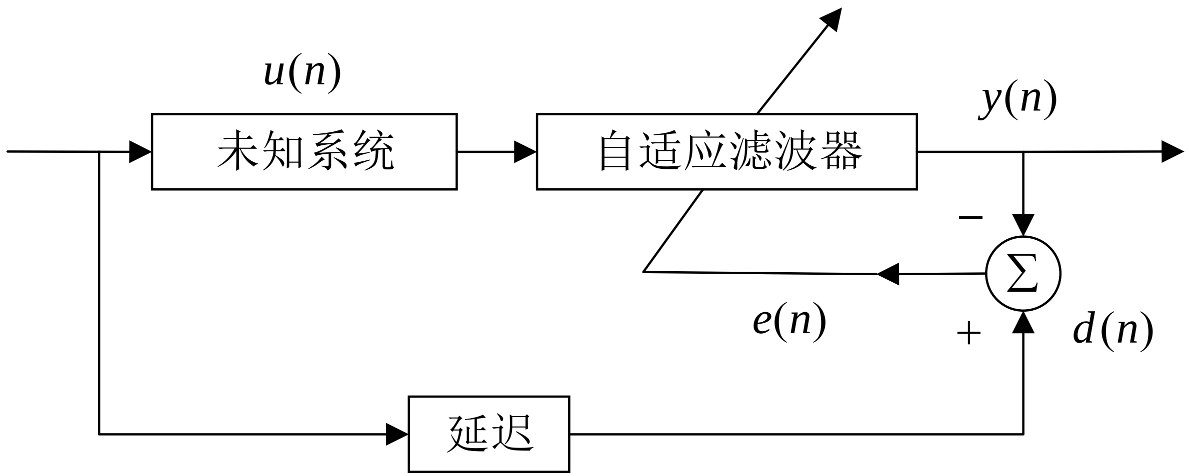
§4.6 自适应滤波器的应用

➢ 系统辨识



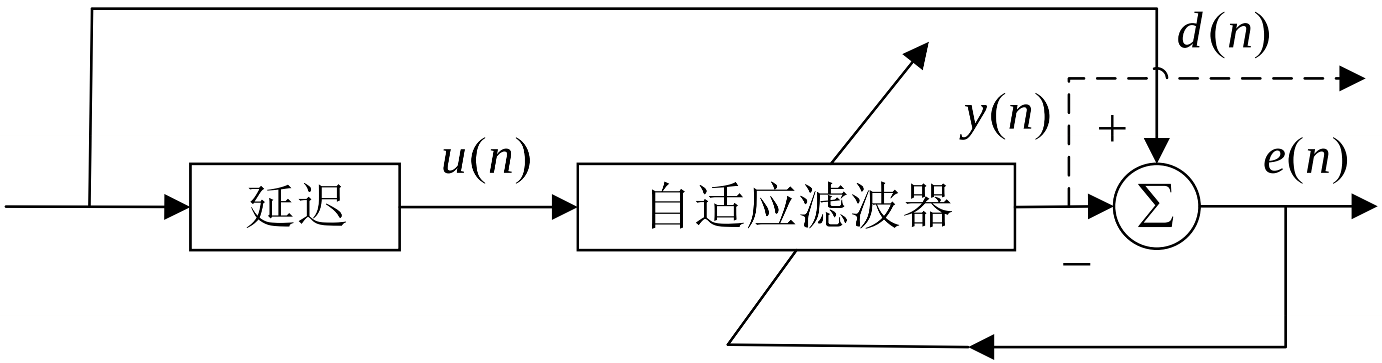
对于系统辨识，自适应滤波器的主要作用是提供逼近未知系统的线性模型。图中 n 代表时刻，$u(n)$ 是系统输入信号，通过未知系统得到期望响应 $d(n)$，通过自适应滤波器得到系统输出 $y(n)$。将期望响应和系统输出相减可得系统误差 $e(n)$。通过将系统误差 $e(n)$ 反馈给自适应滤波器，自适应滤波器可以不断地自适应调节，直到滤波器的参数无限逼近未知系统的参数，最后完成系统辨识。

逆建模



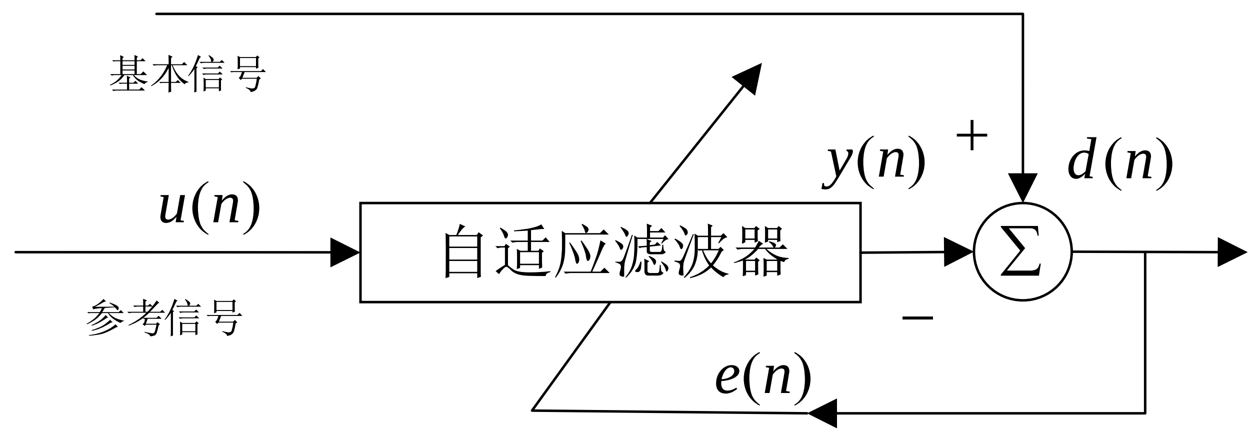
对于信道均衡，自适应滤波器的作用是提供均衡器的一个模型。图中，$u(n)$ 是经过未知系统后的输入，经过自适应滤波器得到系统输出 $y(n)$，系统输入通过延迟器得到含噪的期望输出 $d(n)$，用期望输出与系统输出之差 $e(n)$ 反馈调节自适应滤波器，直至自适应滤波器逼近未知系统的逆模型。

➢ 信号预测



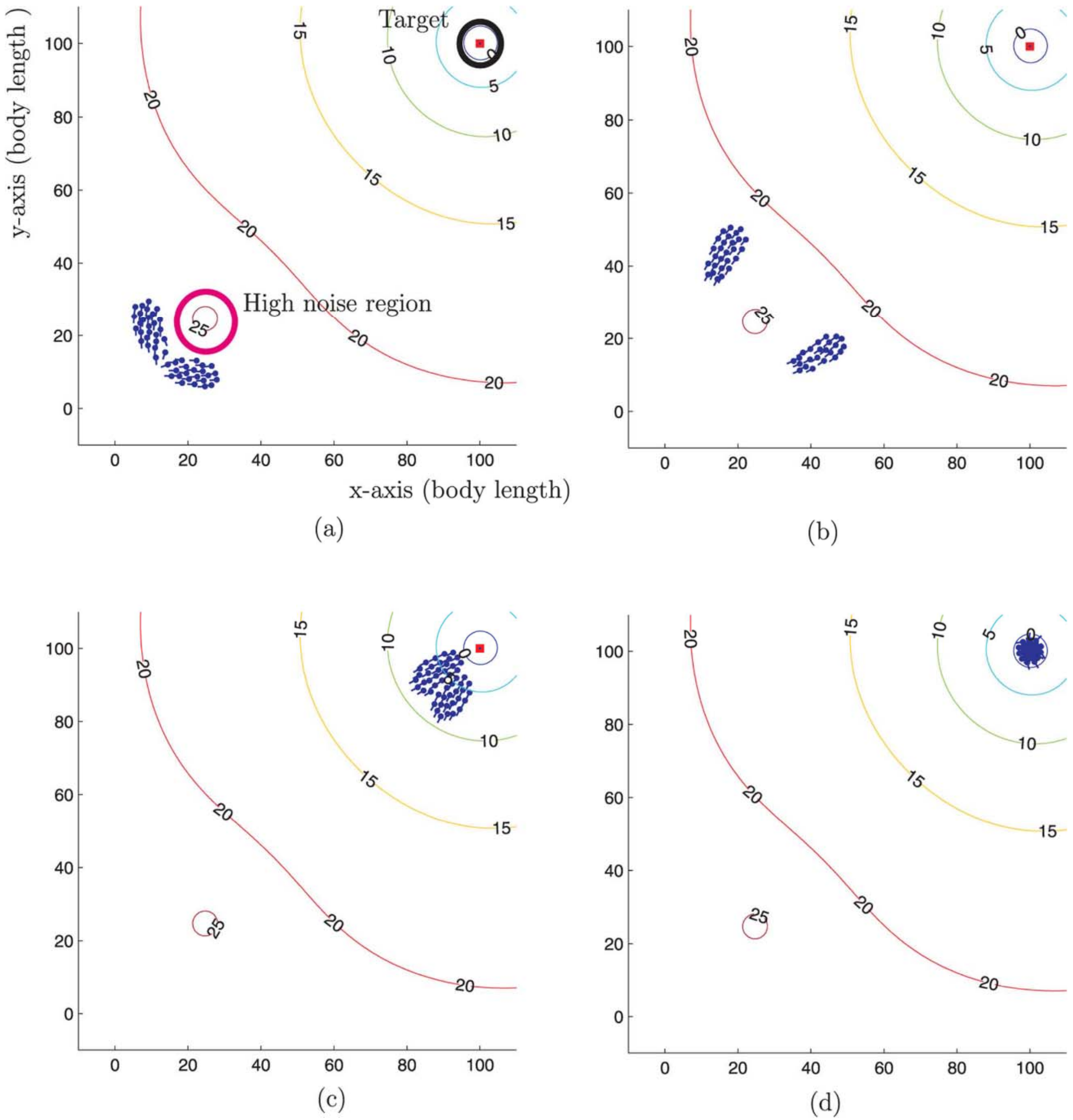
对于信号预测，自适应滤波器的作用就是对随机信号的当前值提供一个最佳的预测。通过延迟器后的系统输入$u(n)$经过自适应滤波器后得到系统输出$y(n)$，当前的信号值$d(n)$作为期望响应，两者之差$e(n)$反馈到自适应滤波器，自适应学习后的系统输出可以作为下个时刻信号值的预测。

➢ 噪声消除



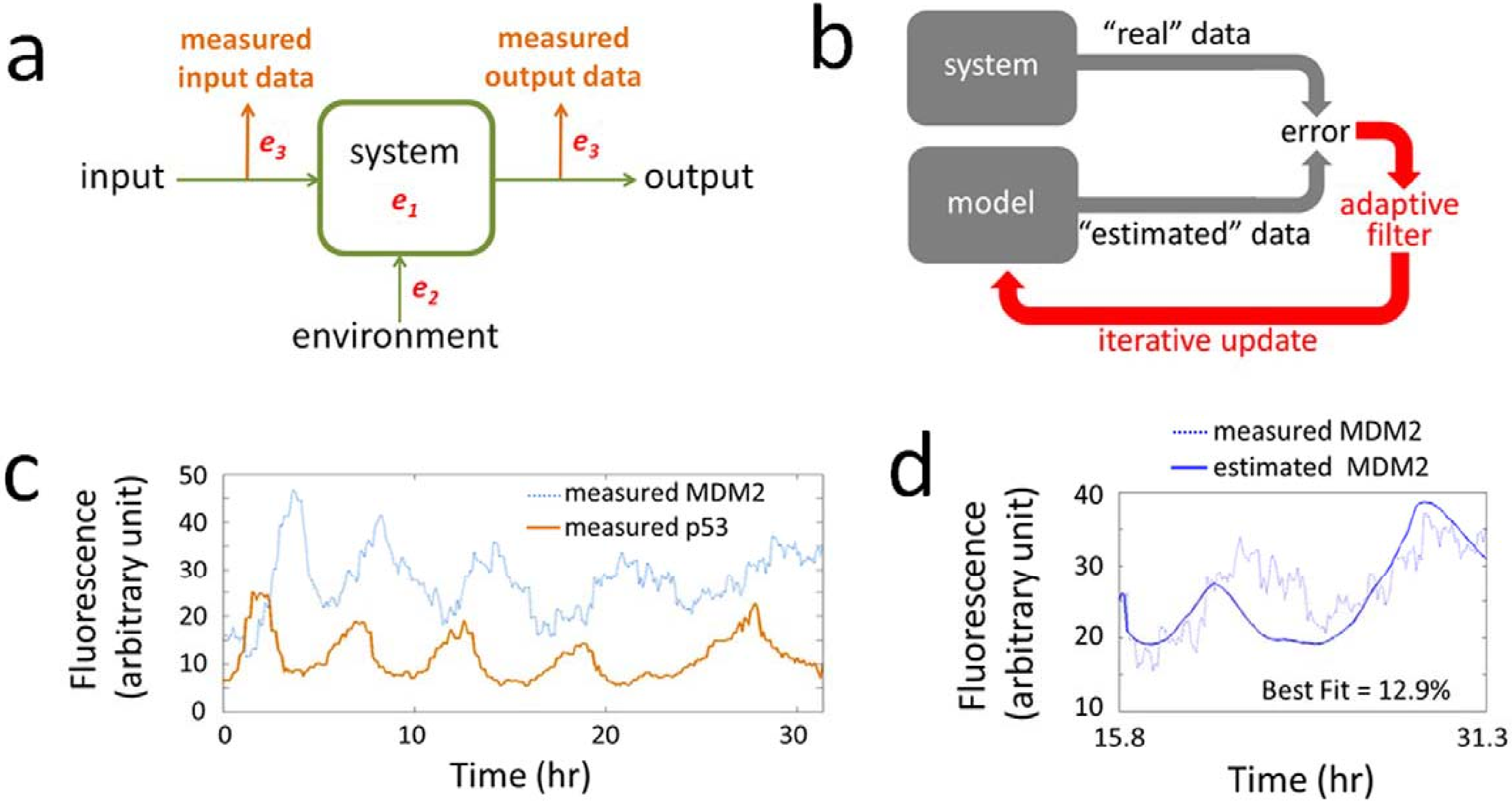
对于噪声消除，自适应滤波器主要作用是消除信号中无用的噪声干扰。图中，$u(n)$ 是系统输入，经过自适应滤波器滤波得到系统输出 $y(n)$，此时将给定的期望响应 $d(n)$ 和系统输出相减得到系统误差 $e(n)$，自适应滤波器利用系统误差自适应反馈调节，直到系统误差小于预定值，从而达到消除噪声干扰的目的。

➤ 目标定位与跟踪



(a) (b) (c) (d)

基因网络建模



a
measured input data
measured output data
input
system
e_1
e_2
e_3
e_3
output
environment
b
system
"real" data
error
model
"estimated" data
adaptive filter
iterative update
c
Fluorescence (arbitrary unit)
measured MDM2
measured p53
Time (hr)
d
measured MDM2
estimated MDM2
Fluorescence (arbitrary unit)
Best Fit = 12.9%
15.8
31.3
Time (hr)

动物群体运动建模